CLAIMS

What is claimed is:

1. A method of determining the tertiary structure of a macromolecule, comprising the steps of:

imposing physical distance constraints between residues of the macromolecule;

fragmenting the macromolecule into smaller molecular fragments;

subjecting the fragments to an identification procedure; and

analyzing identification information obtained from the identification procedure to provide three-dimensional structural information on the macromolecule.

2. The method of claim 1, wherein the identification procedure comprises mass measurement of the fragments using mass spectrometric analysis.

3. The method of claim 2, wherein the identification procedure comprises sequence identification.

4. The method of claim 1, wherein the analyzing information step comprises:

assigning scoring values to the fragments based on the identification of the fragments;

generating hypothetical structures by comparing the macromolecule to related compounds of known structure; and

evaluating the hypothetical structures by considering the distance constraints.

5. The method of claim 4, further comprising:

conducting homology modeling analysis of hypothetical structures which best fit the distance constraints.

6. The method of claim 1, wherein the macromolecule comprises at least one amino acid.

7. The method of claim 1, wherein the macromolecule comprises RNA or DNA.

8. A method of determining the tertiary structure of a protein, comprising the steps of:

reacting a protein to be analyzed with at least one crosslinking reagent, said reagent comprising at least two reactive groups;

enriching the reaction product for molecules having intramolecular crosslinks;

carrying out proteolysis on the enriched reaction product to yield protein_fragments;

subjecting the protein fragments to peptide identification analysis; and

analyzing information obtained from the peptide identification analysis to provide information on the three dimensional structure of the macromolecule.

9. The method of claim 8, wherein the crosslinking reagent is a bifunctional crosslinker.

10. The method of claim 9, wherein the crosslinking reagent is an amine-specific homobifunctional crosslinker.

11. The method of claim 8, wherein the protein is reacted with a plurality of crosslinking agents having different specificities for reactive sites on the protein.

12. The method of claim 8, wherein the protein is reacted with a plurality of crosslinking reagents having varying lengths between reactive groups.

13. The method of claim 1, wherein the reaction with the crosslinker is optimized to produce an average number of one crosslinker modification per macromolecule.

14. The method of claim 8, wherein the reaction product is enriched for molecules having intramolecular crosslinks by physical removal of proteins having intermolecular crosslinks.

15. The method of claim 14, wherein the molecules having intermolecular links are removed using size exclusion chromatography.

16. The method of claim 8, wherein the peptide identification analysis is comprised of chromatography and mass spectrometric analysis.

17. The method of claim 16, wherein the chromatography is reverse-phase separation using C4, C8 and C18 separation schemes.

18. The method of claim 16, wherein the mass spectrometric analysis is carried out using matrix-assisted laser desorption ionization (MALDI) time-of-flight (TOF) instrumentation or electrospray ionization (ESI) time-of-flight (TOF) instruments.

19. The method of claim 8, wherein the peptide identification analysis is comprised of peptide sequencing.

20. The method of claim 8, wherein the analyzing information step comprises:

assigning values to the proteolyzed products based on mass spectrometric data;

generating hypothetical structures by comparing the macromolecule to related compounds of known structure; and

evaluating the hypothetical structures by considering distance constraints obtained from crosslinking data.

21. The method of claim 20, further comprising:

conducting homology modeling analysis of hypothetical structures which fit the distance constraints.

22. The method of claim 20, wherein assigning values is carried out by constructing a virtual library of proteolyzed products which library is indexed by a criteria selected from the group consisting of monoisotopic data and average mass data.

23. The method of claim 20, wherein hypothetical structures are generated using a threading program for fold prediction.

24. The method of claim 20, wherein the hypothetical structures are generated with the use of an equation

$$E_t = \sum_{j=0}^{j<=i} 0 \text{ if } d_j <= d_o, \quad d_j - d_o \text{ if } d_j > 0$$

wherein E_t is the total constraint error, d_o is the pairwise distance separation, d_i is the pairwise

distance defined by the structure by constraint *j* and *i* is the total number of distance constraints.

25. The method of claim 21, wherein homology modeling analysis is carried out using a threading alignment to match components of the macromolecule to spatial positions of those components in a structure of the macromolecule.

26. A method of determining information regarding a three dimensional structure of an amino acid sequence, comprising the steps of:

reacting an amino acid sequence with a crosslinking reagent comprised of two reactive groups and a detectable label to obtain a reaction product, wherein the reactive groups are separated by a distance of from about 5Å to about 20 Å;

subjecting the reaction product to size separation using chromatography;

carrying out proteolysis on a size separated portion of the reaction product and isolating away a portion of reaction product which remains bound to a detectable label of the crosslinking reagent;

performing mass spectrometric analysis on the isolated portion of reaction product comprising detectable labels;

computing the mass of possible reaction products and comparing such to actual experimental masses to provide information relating to a three dimensional structure of the amino acid sequence.

27. A system for determining structural details of a molecule, the system comprising:

a mass spectrometer; and

a computational system that accepts input data from the mass spectrometer, the input data comprising mass information from actual fragments of the molecule, wherein the molecule has had at least one distance constraint imposed on it, and wherein the computational system outputs the structural details of the molecule after matching the input data with expected fragments of the molecule that have been generated or stored by the computational system.

28. The system of claim 27 wherein the molecule is a polypeptide.

29. The system of claim 27 wherein the molecule is a nucleic acid.

30. The system of claim 27 wherein the distance constraint is imposed by a polypeptide cross-linker.

31. The system of claim 27 wherein the distance constraint is imposed by BS3.

32. The system of claim 27 wherein the number of distance constraints is less than about 20.

33. The system of claim 27 wherein the molecule is a polypeptide, and the number of distance constraints is less than about 20% of the number of the number of amino acid residues in the polypeptide.

34. The system of claim 27 wherein the input data is from a mass spectrometer.

35. The system of claim 27 wherein the mass spectrometer is a MALDI or ESI mass spectrometer.

36. The system of claim 27 wherein candidate structures for the molecule are generated by constrained threading of a primary sequence through a known protein fold.

37. The system of claim 27 wherein the structural details of the molecule comprise tertiary structure information.

38. The system of claim 27 wherein the structural details of the molecule comprise a three-dimensional coordinate map.

39. The system of claim 38 wherein the three-dimensional coordinate map is determined to within about 2 Å to about 5 Å RMS of the actual location of each of the atoms of the molecule.

40. The system of claim 27 wherein the structural details of the molecule are generated using homology modeling.

41. A computer system for determining structural details of a molecule, the computer system comprising:

one or more processors; and

one or more user input devices;

wherein the computer system accepts input data, the input data comprising mass information from actual fragments of the molecule, wherein the molecule has had at least one distance constraint imposed on it, and wherein the computer system outputs structural details of the molecule after comparing the input data with expected fragments of the molecule.

42. The system of claim 41 wherein the number of distance constraints is less than about 20.

43. The system of claim 41 wherein the molecule is a polypeptide, and the number of distance constraints is less than about 20% of the number of the number of amino acid residues in the polypeptide.

44. The system of claim 41 wherein candidate structures for the molecule are generated by constrained threading of a primary sequence through a known protein fold.

45. The system of claim 41 wherein the structural details of the molecule comprise tertiary structure information.

46. The system of claim 41 wherein the structural details of the molecule comprise a three-dimensional coordinate map.

47. The system of claim 46 wherein the three-dimensional coordinate map is determined to within 2 Å to about 5 Å RMS of the actual location of each of the atoms of the molecule.

48. The system of claim 41 wherein the structural details of the molecule are generated using homology modeling.

49. A method implemented on a computer system for scoring candidate structures of a molecule, the method comprising the steps of:

accepting input data, the input data comprising mass information from actual fragments of the molecule, wherein the molecule has at least one distance constraint imposed on it;

generating or storing expected fragments of the molecule;

matching the mass information to the expected fragments of the molecule to generate distance constraint information; and

scoring the candidate structures based on how well they fit the distance constraint information.

50. The method of claim 49 wherein the molecule is a polypeptide.

51. The method of claim 49 wherein the molecule is a nucleic acid.

52. The method of claim 49 wherein the distance constraint is imposed by a polypeptide cross-linker.

53. The method of claim 49 wherein the distance constraint is imposed by BS3.

54. The method of claim 49 wherein the number of distance constraints is less than about 20.

55. The method of claim 49 wherein the molecule is a polypeptide, and the number of distance constraints is less than about 20% of the number of the number of amino acid residues in the polypeptide.

56. The method of claim 49 wherein the input data is from a mass spectrometer.

57. The method of claim 49 wherein the mass spectrometer is a MALDI or ESI mass spectrometer.

58. The method of claim 49 wherein the candidate structures are generated by constrained threading of a primary sequence through a known protein fold.

59. The method of claim 49 wherein the candidate structures are determined to a secondary structure level.

60. The method of claim 49 further comprising generating and outputting structural details of the molecule.

61. The method of claim 49 wherein the structural details of the molecule comprise tertiary structure information.

62. The method of claim 49 wherein the structural details of the molecule comprise a three-dimensional coordinate map.

63. The method of claim 63 wherein the three-dimensional coordinate map is determined to within about 2 Å to about 5 Å RMS of the actual location of each of the atoms of the molecule.

64. The method of claim 49 wherein the structural details of the molecule are generated using homology modeling.

65. A computer-program product comprising a computer-readable medium and program instructions provided via the computer-readable medium, the program instructions comprising instructions for scoring candidate structures of a molecule, the instructions specifying:

accepting input data, the input data comprising mass information from actual fragments of the molecule, wherein the molecule has at least one distance constraint imposed on it;

generating or storing expected fragments of the molecule;

matching the mass information to the expected fragments of the molecule to generate distance constraint information; and

scoring the candidate structures based on how well they fit the distance constraint information.

66. The computer-program product of claim 65 wherein the number of distance constraints is less than about 20.

67. The computer-program product of claim 65 wherein the molecule is a polypeptide, and the number of distance constraints is less than about 20% of the number of the number of amino acid residues in the polypeptide.

68. The computer-program product of claim 65 wherein the candidate structures are generated by constrained threading of a primary sequence through a known protein fold.

69. The computer-program product of claim 65 further comprising generating and outputting structural details of the molecule.

70. The computer-program product of claim 65 wherein the structural details of the molecule comprise a three-dimensional coordinate map.

71. The computer-program product of claim 70 wherein the three-dimensional coordinate map is determined to within about 2 Å to about 5 Å RMS of the actual location of each of the atoms of the molecule.

72. The computer-program product of claim 65 wherein the structural details of the molecule are generated using homology modeling.

73. The method of claim 19 wherein the peptide sequencing is comprised of Edman sequencing.

74. The method of claim 21, further comprising:

choosing hypothetical structures which best fit the distance constraints.